



Stochastic Tunneling of Two Mutations in a Population of Cancer Cells

Citation

Haeno, Hiroshi, Yosef E. Maruvka, Yoh Iwasa, and Franziska Michor. 2013. "Stochastic Tunneling of Two Mutations in a Population of Cancer Cells." PLoS ONE 8 (6): e65724. doi:10.1371/journal.pone.0065724. <http://dx.doi.org/10.1371/journal.pone.0065724>.

Published Version

doi:10.1371/journal.pone.0065724

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11717543>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Stochastic Tunneling of Two Mutations in a Population of Cancer Cells

Hiroshi Haeno¹, Yosef E. Maruvka², Yoh Iwasa^{1*}, Franziska Michor^{2*}

1 Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka, Japan, **2** Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America

Abstract

Cancer initiation, progression, and the emergence of drug resistance are driven by specific genetic and/or epigenetic alterations such as point mutations, structural alterations, DNA methylation and histone modification changes. These alterations may confer advantageous, deleterious or neutral effects to mutated cells. Previous studies showed that cells harboring two particular alterations may arise in a fixed-size population even in the absence of an intermediate state in which cells harboring only the first alteration take over the population; this phenomenon is called stochastic tunneling. Here, we investigated a stochastic Moran model in which two alterations emerge in a cell population of fixed size. We developed a novel approach to comprehensively describe the evolutionary dynamics of stochastic tunneling of two mutations. We considered the scenarios of large mutation rates and various fitness values and validated the accuracy of the mathematical predictions with exact stochastic computer simulations. Our theory is applicable to situations in which two alterations are accumulated in a fixed-size population of binary dividing cells.

Citation: Haeno H, Maruvka YE, Iwasa Y, Michor F (2013) Stochastic Tunneling of Two Mutations in a Population of Cancer Cells. PLoS ONE 8(6): e65724. doi:10.1371/journal.pone.0065724

Editor: Frank Emmert-Streib, Queen's University Belfast, United Kingdom

Received: December 19, 2012; **Accepted:** April 26, 2013; **Published:** June 26, 2013

Copyright: © 2013 Haeno et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by NCI. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: michor@jimmy.harvard.edu (FM); yohiwasa@kyudai.jp (YI)

† These authors contributed equally to this work.

Introduction

Genetic and epigenetic alterations in signaling pathways, DNA repair mechanisms, the cell cycle, and apoptosis lead to abnormal reproduction, death, migration, genome stability, and other behaviors of cells, which may lead to the onset and progression of cancer [1]. For example, homozygous inactivation of the RB1 gene causes the childhood eye cancer retinoblastoma [2]. Similarly, a reciprocal translocation between chromosomes 9 and 22 leads to the creation of the BCR-ABL fusion oncoprotein resulting in chronic myeloid leukemia [3,4]. Epigenetic alterations can also induce abnormalities in gene expression within cancer cells [5]. Furthermore, drug resistance in cancer cells is acquired by genetic and/or epigenetic changes: in the treatment of chronic myeloid leukemia, for instance, combination therapy of imatinib (Gleevec, STI571) and dasatinib (BMS-35482) often fails due to the emergence of only one or two genetic alterations within the tyrosine kinase domain of BCR-ABL [6].

While experimental studies have identified specific (epi)genetic changes and their consequences for cancer progression and drug resistance, mathematical investigations have provided insights into how tumor cells accumulate such alterations during tumorigenesis. In the 1950s, the multi-stage theory of carcinogenesis was proposed when Nordling, Armitage and Doll, and Fisher investigated the age distribution of cancer incidence with mathematical approaches [7,8,9]. In 1971, Knudson revealed, utilizing statistical analyses of the retinoblastoma incidence data, that two hits in an “anti-oncogene” are the rate-limiting steps in

this disease [2]; this gene was later identified as the tumor suppressor RB1 [10]. In recent years, biological knowledge about population dynamics and molecular mechanisms of tumorigenesis, invasion, and therapeutic resistance have been incorporated into the mathematical models; for instance, tissue structures in particular cancer types [11,12,13,14,15,16] and the evolution of drug resistance in cancer cells [17,18,19] were considered.

Much effort has been devoted to elucidating the dynamics of accumulating two (epi)genetic alterations in a population of a fixed number of cells. The theory that reveals the dynamics of accumulation of two specific mutations in a population is useful for predicting the risk of emergence and the rate of progression of cancer cells, and also for the kinetics of drug resistance. Moreover, the theory can be extended to more complicated cases in which more than two specific mutations play a role in malignant lesions. In 2003, Komarova et al. [20] derived analytic solutions of stochastic mutation-selection networks with an assumption that most of the time, the cell population is homogeneous with respect to relevant mutations. They defined stochastic tunneling as the case in which cells with two mutations appear from a lineage of cells harboring a single mutation; the latter eventually goes extinct instead of reaching fixation. They performed a precise analysis of the existence of stochastic tunnels and explicitly calculated the rate of tunneling [20]. In 2004, Nowak et al. [21] calculated the probability as function of time that at least one cell with two inactivated alleles of a tumor suppressor gene has been generated. They found three different kinetic laws: in small, intermediate, and large populations, it took, respectively, two, one, and zero rate-

limiting steps to inactivate a tumor suppressor. They also studied the effect of chromosomal and other genetic instabilities. Small lesions without genetic instability required a very long time to inactivate the next TSG, whereas the same lesions with genetic instability posed a much greater risk for cancer progression [21]. Iwasa et al. [22], in the same year, derived the explicit tunneling rate for situations in which cells with one mutation were neutral or disadvantageous as compared to wild type cells, with cells with two mutations having the largest fitness. The analytical solutions provided an excellent fit to exact stochastic computer simulations [22]. In 2005, Weinreich and Chao [23] developed an analytical expression for the critical population size that defines the boundary between the regime of sequential fixation of two mutations and that of simultaneous fixation in a Wright-Fisher model; they also investigated the effect of recombination on this phenomenon [23]. In 2008, Schweinsberg investigated the waiting time for a large number of mutations to arise when the fitness change conferred by each mutation is negligible; i.e. when the mutations are neutral [24]. Lynch studied the mean time to fixation of two mutations and the effects of recombination on this process in a large range of population sizes [25]. Weissman et al. [26] and Altland et al. [27] analyzed how recombination affects the expected time to achieve fixation of two mutations under the assumption that intermediate cell types are disadvantageous.

In 2009, Weissman et al. [28] calculated the rate of stochastic tunneling as a function of the mutation rates, the population size, and the fitness of the intermediate population harboring only a single mutation in the Wright-Fisher model. They found that when intermediate populations were close to neutral as compared to wild type cells, then stochastic tunneling easily emerged in large populations. In small populations, however, stochastic tunneling was much less likely to arise [28]. Later on, Proulx used elementary methods of analyzing stochastic processes to derive the probability of tunneling in the limit of large population sizes for both the Moran and Wright-Fisher models. He found that the probability of stochastic tunneling was twice as large in the Wright-Fisher model as in the Moran model [29].

Finally, diffusion approximations also represent a useful method for describing the evolutionary process of accumulating mutations in a large population of cells under the assumption of weak selection [30]. In 2009, Lehmann and Rousset [31] investigated multi-locus fixation probabilities under arbitrary strengths of selection in the Wright-Fisher model by using the tools of diffusion approximations. They showed that such fixation probabilities could be expressed in terms of selection coefficients weighted by the mean first passages times of ancestral gene lineages within a single ancestor. They then applied these results to investigate the Hill-Robertson Interference, i.e. stochastic tunneling of cell lineages [31].

Despite a wealth of forays into the dynamics of stochastic tunneling of two mutations within populations of cells, several critical questions remain. For instance, currently available approaches do not provide accurate predictions for situations in which mutation rates are large. Such scenarios, however, are important when considering mutation accumulation in cancer cells since many tumor types exhibit mutator phenotypes [32–37]. Furthermore, existing methods do not take into account all possible fitness effects of the individual cell types – such as increased fitness of cells with one mutation as compared to those with zero or two mutations.

In this paper, we addressed these scenarios to provide a general description of stochastic tunneling in a tumor cell population of constant size. Such a model describes many situations arising during tumorigenesis such as the dynamics of cancer initiation

from a cellular compartment of a healthy tissue as well as the chronic phase of tumor progression [21,38]. We designed three methods to calculate the probability of existence of a homogeneous population of cells, all of which harbor two mutations, at an arbitrary time point. One method demonstrated an accurate fit against all scenarios in numerical simulations, but had a large computational cost. The second method showed a very good fit with small computational cost; however, the predictions were not accurate in cases in which cells with two mutations had the same fitness as wild type cells. The last method produced accurate results in the latter situation of neutral fitness. By utilizing the best method for each parameter condition, we obtained an accurate approximation for the probability of a homogeneous population of cells with two mutations over time.

Methods

The mathematical model

Let us consider a population of N reproducing cells proliferating according to the Moran process [39]. One elementary time step of this process consists of a cell division and a cell death. For each division event, a cell is chosen at random proportional to fitness; the division event may produce a mutated daughter cell with a small probability. For each death event, one cell is chosen at random from the population. The total number of cells, N , is constant over time. These cells may accumulate (epi)genetic alterations and/or structural genomic changes; these are collectively referred to as “mutations”. We consider three types of cells: those harboring no mutations, denoted as type-0 cells, those harboring the first of a sequence of two mutations, denoted as type-1 cells, and those harboring both mutations, denoted as type-2 cells. Initially, the population consists entirely of type-0 cells; these cells have relative fitness (i.e. growth rate) r_0 . During each type-0 cell division, a type-1 cell may arise with probability equal to the mutation rate u_1 . The fitness of type-1 cells is given by r_1 . Finally, a type-2 cell may arise with probability u_2 per type-1 cell division and has fitness r_2 . We assume that there is no back mutation because a mutation that exactly reverses the functional change caused by a specific mutation is rare compared to a mutation that causes a phenotypic change. Time is measured in units of cell divisions. Eventually, a type-2 cells will appear and may become dominant in the population; this event represents the evolution of adaptive cells.

In previous studies [20,22], three states of a homogeneous population were considered: states in which all cells in the population are of type-0, type-1 or type-2 (**Figure 1a**). The authors then approximated the dynamics of fixation and tunneling in a heterogeneous population by using a fixation probability and a tunneling rate. This approximation, however, neglects the time from the appearance of a mutated cell to its fixation, as well as the effects of any additional mutational events during the time until fixation; this choice was made due to the observation that the waiting time of new mutation is usually much longer than the time of fixation in the parameter regimes considered. In some situations arising during tumorigenesis, however, these effects cannot be neglected – especially when mutation rates are large. In those cases, the previously derived approximation does not provide an accurate fit to the exact solution of the system. We thus aimed to consider the evolutionary dynamics of two mutations arising in a heterogeneous population using the methods described in the following (**Figure 1b**).

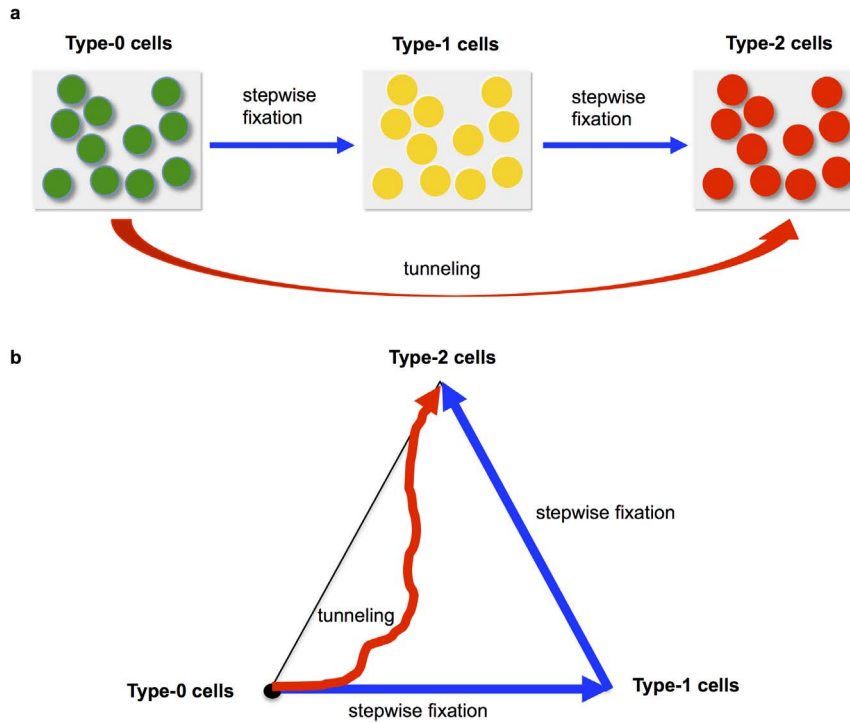


Figure 1. Schematic illustration of the model. Panel a shows the previously published approach to describing the evolutionary dynamics of two mutations in a fixed-size population of cells; only the transitions between homogeneous populations are considered. Panel b displays our novel approach, which encompasses considering the transitions in a heterogeneous population in detail.
doi:10.1371/journal.pone.0065724.g001

Monte-Carlo simulations

We first performed Monte-Carlos simulations of the model describe above. Denote the number of type-0, type-1, and type-2 cells by n_0 , n_1 , and n_2 , respectively. Time is measured in cell cycles. During each time unit, one cell division and one cell death event occur to maintain a constant total number of cells. During one time step, the probability of a cell division of each cell type is given by

$$\begin{aligned}\Pr[(n_0, n_1, n_2) \rightarrow (n_0 + 1, n_1, n_2)] &= \frac{r_0 n_0 (1 - u_1)}{r_0 n_0 + r_1 n_1 + r_2 n_2} \\ \Pr[(n_0, n_1, n_2) \rightarrow (n_0, n_1 + 1, n_2)] &= \frac{r_0 n_0 u_1 + r_1 n_1 (1 - u_2)}{r_0 n_0 + r_1 n_1 + r_2 n_2}, \\ \Pr[(n_0, n_1, n_2) \rightarrow (n_0, n_1, n_2 + 1)] &= \frac{r_1 n_1 u_2 + r_2 n_2}{r_0 n_0 + r_1 n_1 + r_2 n_2}\end{aligned}$$

while the probability of a cell death of each cell type is given by

$$\begin{aligned}\Pr[(n_0, n_1, n_2) \rightarrow (n_0 - 1, n_1, n_2)] &= \frac{n_0}{N} \\ \Pr[(n_0, n_1, n_2) \rightarrow (n_0, n_1 - 1, n_2)] &= \frac{n_1}{N} \\ \Pr[(n_0, n_1, n_2) \rightarrow (n_0, n_1, n_2 - 1)] &= \frac{n_2}{N}\end{aligned}$$

The initial condition is given by $n_0 = N$ and $n_1 = n_2 = 0$. We performed 100,000 runs for each parameter set and obtained the

fraction of cases in which the population consists entirely of type-2 cells at a given time.

A novel approach

We extended our previously obtained results [22] to accurately describe situations in which mutation rates are large by considering the detailed transitions between states within a heterogeneous population. Denote by $X_0(t)$, $X_1(t)$, and $X_2(t)$, respectively, the probabilities at time t that the system consists exclusively of type-0, type-1, and type-2 cells. Then the dynamics of the population can be described by the forward Kolmogorov differential equations:

$$\frac{dX_0}{dt} = -aX_0 - bX_0, \quad (1a)$$

$$\frac{dX_1}{dt} = aX_0 - cX_1, \quad (1b)$$

$$\frac{dX_2}{dt} = cX_1 + bX_0. \quad (1c)$$

The rate at which the population transitions from type-0 to type-1, a , is given by

$$a = Nu_1\rho_1. \quad (2)$$

Here ρ_1 denotes the fixation probability of one type-1 cell in a population of $N-1$ type-0 cells and given by

$$\rho_1 = \frac{(1 - r_0(1 - u_1)/(r_1 + r_0u_1))}{\left(1 - (r_0(1 - u_1)/(r_1 + r_0u_1))^N\right)}. \quad (3)$$

We have included the effect of the mutation rate in the fixation probability because, in situations when u_1 is very large, additional mutations can occur during the fixation of the former lineage. If $u_1 = 0$, then $\rho_1 = (1 - r_0/r_1) / \left(1 - (r_0/r_1)^N\right)$, which was derived previously [20].

The tunneling rate, i.e. the rate at which the population transitions from type-0 to type-2 without the fixation of type-1 cells, b , is given by

$$b = Nu_1(1 - V_1 - \rho_1). \quad (4)$$

Here V_i denotes the probability of non-appearance or extinction of a new type-2 lineage from i type-1 cells. With $V_0 = 1$ and $V_N = 0$, V_i can numerically be calculated from the following equation:

$$\frac{-r_1u_2\rho_3V_i + [r_1(1 - u_2)(V_{i+1} - V_i) + r_0(V_{i-1} - V_i)]}{ir_1(1 - u_2) + r_0(N - i)} = 0. \quad (5)$$

Here $\rho_3 = (1 - r_0(1 - u_1)/(r_2 + r_1u_2)) / \left(1 - (r_0(1 - u_1)/(r_2 + r_1u_2))^N\right)$. In both equations of V_i and ρ_3 , we include mutational events, which may increase or decrease the relative fitness of each cell type. See [22] for a detailed derivation of V_i .

Next, let us consider the following quantity:

$$G(t) = \Pr \left(\begin{array}{c} \text{type - 2 cells are fixed at time } t \\ \text{type - 1 cells are fixed in the} \\ \text{initial population} \end{array} \right).$$

Then we have

$$X_2(t) = \int_0^t bX_0(s)ds + \int_0^t G(t-s)aX_0(s)ds. \quad (6)$$

If we assume

$$G(t) = 1 - e^{-ct}, \quad (7)$$

where $c = Nu_2(1 - r_1/r_2) / (1 - (r_1/r_2)^N)$, then we have

$$X_1(t) = \int_0^t e^{-c(t-s)}aX_0(s)ds. \quad (8)$$

By taking the derivative of Eq. (6) and (8), we obtain Eq. (1). Equation 1 no longer holds, however, when the second mutation rate, u_2 , is very large since Equation 7 does not hold. Therefore, let us next calculate $G(t)$ in a heterogeneous population of type-1 and type-2 cells.

Consider the $N+1$ states that are classified by the number of type-2 cells, $k=0, 1, 2, \dots, N$. Since we are interested in the situation after the emergence of type-1 cells, the number of type-1 cells becomes $N-k$. Then the transition probabilities are given by

$$\Pr(k \rightarrow k+1, \text{ in } \Delta t) = (N-k)\Delta t \frac{(N-k)r_1u_2 + r_2k}{(N-k)r_1 + r_2k} = \lambda_k \Delta t, \quad (9a)$$

$$\Pr(k \rightarrow k-1, \text{ in } \Delta t) = k \Delta t \frac{(N-k)r_1(1-u_2)}{(N-k)r_1 + r_2k} = \mu_k \Delta t, \quad (9b)$$

$$\Pr(k \rightarrow k, \text{ in } \Delta t) = 1 - \Pr(k \rightarrow k+1, \text{ in } \Delta t) - \Pr(k \rightarrow k-1, \text{ in } \Delta t), \quad (9c) \\ = 1 - \lambda_k \Delta t - \mu_k \Delta t$$

for $k=1, 2, \dots, N-1$. For $k=0$, we have $\mu_0=0$. Note that the transition probability includes the second mutation rate, u_2 , which is normally neglected when deriving the fixation probability in the Moran process due to the assumption of a very small mutation rate. Then we consider the following quantities:

$$q_k(t) = \Pr(\text{type - 2 cells are fixed at time } t | \text{initial number of type - 2 cells is } k), \quad (10)$$

where $k=0, 1, 2, \dots, N$. Hence we have

$$G(t) = q_0(t). \quad (11)$$

By definition, we have the boundary condition, $q_N(t) = 1$, and the initial condition, $q_k(0) = 0$ for $k=1, 2, 3, \dots, N-1$. Then we obtain the following backward equation:

$$q_k(t + \Delta t) = (1 - \lambda_k \Delta t - \mu_k \Delta t)q_k(t) + \lambda_k \Delta t q_{k+1}(t) + \mu_k \Delta t q_{k-1}(t). \quad (12)$$

By taking the limit when $\Delta t \rightarrow 0$, we have

$$\frac{dq_k}{dt} = \lambda_k q_{k+1}(t) - (\lambda_k + \mu_k)q_k(t) + \mu_k q_{k-1}(t) \quad (13) \\ \text{for } k=0, 1, \dots, N-1.$$

Note that from Eq. (1a) and $X_0(0) = 1$, we have $X_0(t) = \exp[-(a+b)t]$. We set the second term of Eq. (6) as

$$L(t) = \int_0^t G(t-s)ae^{-(a+b)s}ds = \int_0^t G(z)ae^{-(a+b)(t-z)}dz. \quad (14)$$

Here $L(0) = 0$ since $G(0) = 0$. Finally, we have

$$X_2(t) = \int_0^t be^{-(a+b)s}ds + L(t) = \frac{b}{a+b}(1 - e^{-(a+b)t}) + L(t). \quad (15)$$

By calculating the derivative of Eq. (14) we have

$$\begin{aligned}\frac{dL}{dt} &= aG(t) - (a+b)L \\ &= aq_0(t) - (a+b)L.\end{aligned}\quad (16)$$

Eq. (15) provides good predictions for all ranges of mutation rates and relative fitness values of mutated cells, except when type-0 and type-2 cells are neutral ($r_0 = r_2$) and the relative fitness of type-2 cells is smaller than that of type-0 cells (**Figure S2**). Although this method works in a wide parameter region, in order to investigate parameter regions where it does not accurately predict the exact dynamics, we consider two alternative methods.

Systematic calculation of all transitions

Let us denote by (i, j) the state of the system in which the numbers of type-1 and type-2 cells are i and j , respectively. The state is confined within the following conditions: $i \geq 0$, $j \geq 0$, and $N - i - j \geq 0$. The system will eventually be absorbed into the state $(i, j) = (0, N)$, indicating that type-2 cells have reached fixation (i.e., 100% frequency) within the population. The fixation probability of type-2 cells from each state is then determined by using a backward calculation. For $i = 0, 1, 2, \dots, N$, and $j = 0, 1, 2, 3, \dots, N$, satisfying $i+j \leq N$, we consider the probability, $W(i, j, t)$, that type-2 cells have reached fixation before time t , starting from state (i, j) . The boundary condition is given by

$$W(0, N, t) = 1, \quad (17a)$$

while the initial condition is given by

$$W(i, j, 0) = 0 \text{ for } (i, j) \neq (0, N), \quad (17b)$$

$$W(0, N, 0) = 1. \quad (17c)$$

Let us next consider the state transitions and derive the recurrence formulas for $W(i, j, t)$. Within a short time interval, Δt , there exist six transitions:

[1] A transition from (i, j) to $(i+1, j)$ occurs when a type-0 cell dies and is replaced by a type-1 cell. There are two ways for this to occur: (i) a type-0 cell may die and a type-1 cell may divide (without mutating to give rise to a type-2 cell) or (ii) a type-0 cell may die and a type-0 cell may divide and mutate into a new type-1 cell. Then the transition probability is given by $(N-i-j)\Delta t B[ir_1(1-u_2) + (N-i-j)r_0u_1]$. Here $(N-i-j)\Delta t$ represents the probability of death of a type-0 cell during a short time interval, $B[ir_1(1-u_2) + (N-i-j)r_0u_1]$ represents the probability of increasing the number of type-1 cells, and $B = \frac{1}{(N-i-j)r_0 + ir_1 + jr_2}$ gives the inverse of the total reaction rate.

[2] A transition from (i, j) to $(i-1, j)$ occurs when a type-1 cell dies and is replaced by a type-0 cell. The probability of this event is given by $i\Delta t B(N-i-j)r_0(1-u_1)$.

[3] A transition from (i, j) to $(i, j+1)$ occurs when a type-0 cell dies and either a type-2 cell divides or a type-1 cell divides with a mutation, giving rise to a new type-2 cell. The transition probability of this event is given by $(N-i-j)\Delta t B(jr_2 + ir_1u_2)$.

[4] A transition from (i, j) to $(i, j-1)$ occurs when a type-2 cell dies and is replaced by type-0 cell. This probability is this event given by $j\Delta t B(N-i-j)r_0(1-u_1)$.

[5] A transition from (i, j) to $(i+1, j-1)$ occurs when a type-2 cell dies and either a type-1 cell divides without a mutation or a type-0 cell divides with a mutation. The transition probability for this event is given by $j\Delta t B[ir_1(1-u_2) + (N-i-j)r_0u_1]$.

[6] A transition from (i, j) to $(i-1, j+1)$ occurs when a type-1 cell dies and either a type-2 cell divides or a type-1 cell divides with a mutation. The transition probability for this event is given by $i\Delta t B(jr_2 + ir_1u_2)$.

Furthermore, there is a possibility that no transition occurs during a short time interval; the probability of no event occurring is given by one minus the sum of all the transition probabilities outlined above.

Considering these transitions between states, we have the following recurrence formula:

$$\begin{aligned}W(i, j; t + \Delta t) &= (N-i-j)\Delta t B[ir_1(1-u_2) + (N-i-j)r_0u_1] \\ &W(i+1, j; t) \\ &+ i\Delta t B(N-i-j)r_0(1-u_1)W(i-1, j; t) \\ &+ (N-i-j)\Delta t B(jr_2 + ir_1u_2)W(i, j+1; t) \\ &+ j\Delta t B(N-i-j)r_0(1-u_1)W(i, j-1; t) \\ &+ j\Delta t B[ir_1(1-u_2) + (N-i-j)r_0u_1] \\ &W(i+1, j-1; t) \\ &+ i\Delta t B(jr_2 + ir_1u_2)W(i-1, j+1; t) \\ &+ (1 - [\text{sum of all the transition probabilities}]) \\ &W(i, j; t).\end{aligned}\quad (18)$$

The left hand side of Eq. (18) denotes the fixation probability of a type-2 cell within the time interval Δt , given that the initial state is (i, j) . The right hand side is composed of the paths according to the type of event occurring during the time interval of length Δt . By calculating the limit when $\Delta t \rightarrow 0$, we have

$$\begin{aligned}\frac{dW}{dt}(i, j; t) &= (N-i-j)B[ir_1(1-u_2) + (N-i-j)r_0u_1] \\ &(W(i+1, j; t) - W(i, j; t)) \\ &+ iB(N-i-j)r_0(1-u_1)(W(i-1, j; t) - W(i, j; t)) \\ &+ (N-i-j)B(jr_2 + ir_1u_2) \\ &(W(i, j+1; t) - W(i, j; t)) \\ &+ jB(N-i-j)r_0(1-u_1)(W(i, j-1; t) - W(i, j; t)) \\ &+ jB[ir_1(1-u_2) + (N-i-j)r_0u_1] \\ &(W(i+1, j-1; t) - W(i, j; t)) \\ &+ iB(jr_2 + ir_1u_2)(W(i-1, j+1; t) - W(i, j; t)).\end{aligned}\quad (19)$$

Using the initial condition Eq. (17b) and Eq. (17c), and the boundary condition Eq. (17a), we can numerically determine $W(0, 0; t)$, which represents the fixation probability of type-2 cells until time t in a population starting from N type-0 cells (**Figure S1**). Although this method provides accurate results, the time necessary for the numerical calculation, i.e. the number of equations, increases in a factorial way as the population size

increases; on the other hand, it increases linearly in the first method. Therefore this method is not suitable for the determination of the dynamics in a large population.

A simulation approach for the neutral case ($r_0 = r_2$)

An analytical formula describing the behavior of a system may serve several goals. One important goal is the ability to quickly obtain a prediction of the expected outcomes of a process, without the need for actually performing the process – no matter whether it is an experimental process or a Monte-Carlo simulation representing a large computational burden. This goal can be also achieved by approximating the time-consuming Monte-Carlo simulation by another Monte-Carlo simulation that is much less computationally expensive. Even though the two simulations differ, the faster one may still serve as a good approximation of the slower one. Note that the use of the Wright-Fisher model in this context solely serves to increase the computational speed of our simulation, and is thus meant as an approximation to the Moran model. The Wright-Fisher model was *not* introduced to study an alternative population model, but instead was used as an approximation to the model under investigation (the Moran model) only.

Here we present the use of the tunneling process in the Wright-Fisher framework as an approximation for the tunneling process in the Moran framework. In the Moran framework, every generation is composed of $O(N)$ random steps, while in the Wright-Fisher framework, the number of randomized steps per generation is independent of N . Instead, it depends only on the number of distinct cell types because there is a need only to generate the number of offspring each type will have in the next generation, and this can be done collectively.

We performed the Wright-Fisher Monte-Carlo simulation in the following way. At a given time t the state of the system is described by the vector $\mathbf{n}(t)$, where n_0 is the number of type-0 cells, n_1 is the number of type-1 cells, and n_2 is the number of type-2 cells. At every generation time step, the current population generates the next generation denoted by $[m_0, m_1, m_2]$ from a multinomial distribution, with a probability vector $\vec{p} = [n_0 \cdot r_0, n_1 \cdot r_1, n_2 \cdot r_2] / \sum_i n_i \cdot r_i$. From the new offspring of type-0 cells, a binomially distributed number, with parameters m_0 and u_1 , mutate and become type-1 cells, and from the offspring of type-1 cells, a binomially distributed number, with parameters m_1 and u_2 , mutate and become type-2 cells. The process starts with N_0 cells of type-0 and stops when one cell type reaches fixation or when the process reaches the maximal time. For a given set of parameter values, 100,000 replicates of the Monte-Carlo simulation were performed and the fixation probability was estimated as the fraction of cases in which type-2 cells reached fixation by time t . In order to compare the Wright-Fisher process to the Moran process, the population size N_0 was then rescaled with the standard scaling of dividing by the standard deviation of the number of offspring each individual cell has, which is $\sqrt{2}$ in the Moran process. Thus the population size used in the Wright-Fisher process is $N_{WF} = N/\sqrt{2}$.

Since the first method performs well for the non-neutral case, $r_0 \neq r_2$, we applied the Wright-Fisher approximation only for the neutral case, $r_0 = r_2$. In general, the Wright-Fisher process has a similar fixation probability as the Moran process, and thus it can serve as a good approximation of the Moran model. In situations in which the fixation probability is very small, the difference between the two processes increases, thus rendering this approx-

imation less exact; however, in these situations the approaches outlined above lead to accurate predictions.

Results

We investigated the quality of fit of the approximations to the numerical results of the exact stochastic computer simulations. **Figure 2** displays the fit between the first approximation and Monte-Carlo simulation results in a wide parameter region (**Figure 2**). However, when the fitness value of type-2 cells is the same as that of type-0 cells, this approximation does not provide accurate predictions (**Figure S1**). We consider this parameter region in greater detail later. The comprehensive analysis showed that the probability of type-2 fixation increases when mutation rates are large and the fitness of type-2 cells is large.

Moreover, we found that there exists an optimal value of the fitness of type-1 cells that maximizes the fixation probability of type-2 cells at a given time point. If the fitness of type-2 cells is the same as that of type-0 cells and if the mutation rates are small, then the optimal value for the fitness of type-1 cells becomes 1 (**Figure 2c**). If the first mutation rate is very large, then a disadvantageous effect of the first mutation leads to the highest probability of type-2 fixation (**Figure 2a**). If the second mutation rate is very large, then an advantageous effect of the first mutation results in the highest probability of type-2 fixation (**Figure 2b–c**). If the fitness of type-2 cells is larger than that of type-0 cells, the optimal fitness of type-1 cells is between that of type-0 and type-2 cells in most cases (**Figure 2d–f**). However, when the first mutation rate is very large and the second mutation rate is very small, then a disadvantageous first mutation again leads to the highest probability of type-2 fixation (**Figure 2d**).

Furthermore, when the second mutation rate is very large and the first mutation rate is low, the optimal fitness of type-1 cells becomes even larger than that of type-2 cells (**Figure 2d–f**). Even though the fitness of type-2 cells is expected to be smaller than that of type-0 cells, fixation may still occur when the population size is small (**Figure 2g–i**). When type-2 cells are advantageous compared to type-0 cells, the tendency of the optimal fitness of type-1 cells does not depend on different values of the population size (**Figure 2j–m**). When time increases, then the fixation probability of population with two mutations also increases (data not shown).

We next investigated the predictions of the alternative method, which determines all transitions between states. Using the initial condition Eq. (17b) and Eq. (17c) and the boundary condition Eq. (17a), we numerically determined $W(0,0;t)$, which represents the fixation probability of type-2 cells until time t in a population starting from N type-0 cells. **Figure 3** and **Figure S2** display the fit of $W(0,0;t)$ against results from direct computer simulations of the Moran model in a wide parameter region of small population sizes. The predictions provide an accurate fit to the simulation results.

Furthermore, we performed computational simulations using the Wright-Fisher framework to obtain the approximate results of the Moran model (see alternative method 2 above). **Figure 4** displays the fit between the results of the Wright-Fisher model and those of the Moran model. This method provides accurate predictions for cases in which the fitness of type-2 cells is the same as the fitness of type-0 cells.

We also investigated the parameter regimes in which the “stochastic tunneling” becomes important (**Figure 5**). Comparing the results by the direct simulation to forward Kolmogorov differential equation without tunneling term (Eq. 1 in [22] with

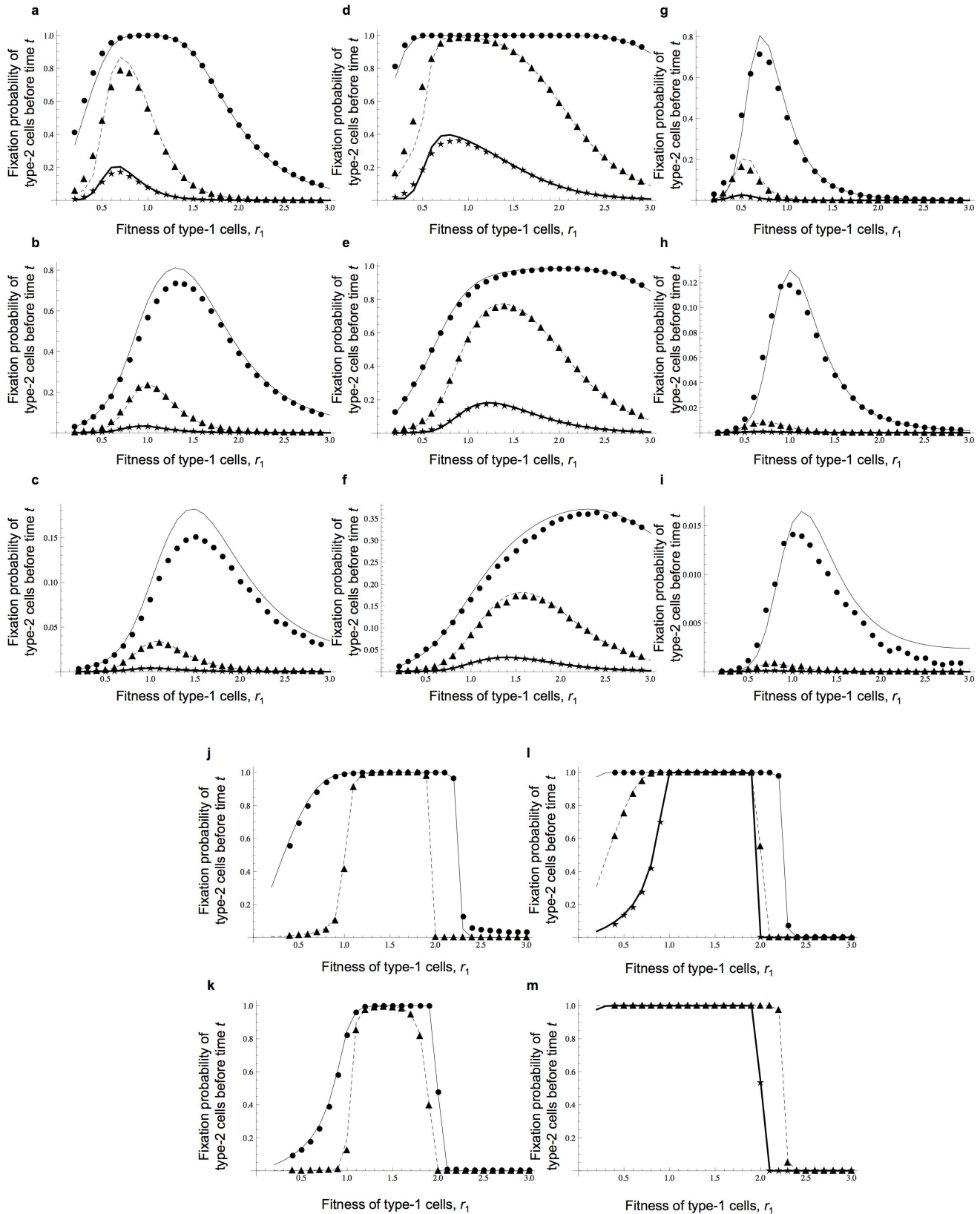


Figure 2. Results of our method. The figure shows the dependence of the probability that type-2 cells are fixed at time t on various parameters. Results by Eq. (15) are indicated by curves and those from direct computer simulations are shown by dots. The results of numerical calculations are connected and shown as a curve. Parameter values are $r_0 = 1$; (a–i) $N = 10$ and $t = 100$; (a–c) $r_2 = 1.0$; (d–f) $r_2 = 2.0$; (g–i) $r_2 = 0.5$; (a), (d), and (g) $u_1 = 0.1$; (b), (e), and (h) $u_1 = 0.01$; and (c), (f), and (i) $u_1 = 0.001$. (a–i) Circles and thin curves represent $u_2 = 0.1$, triangles and dotted lines represent $u_2 = 0.01$, and stars and bold lines represent $u_2 = 0.001$. (j–m) $N = 1000$, $t = 300$ and $r_2 = 2.0$; (j) circles and thin curves represent $u_1 = 0.0001$ and

$u_2 = 0.1$, triangles and dotted lines represent $u_1 = 0.0001$ and $u_2 = 0.001$; (k) circles and thin curves represent $u_1 = 0.0001$ and $u_2 = 0.01$, triangles and dotted lines represent $u_1 = 0.0001$ and $u_2 = 0.0001$; (l) circles and thin curves represent $u_1 = 0.001$ and $u_2 = 0.1$, triangles and dotted lines represent $u_1 = 0.001$ and $u_2 = 0.01$, and stars and bold lines represent $u_1 = 0.001$ and $u_2 = 0.001$; and (m) triangles and dotted lines represent $u_1 = 0.01$ and $u_2 = 0.1$, and stars and bold lines represent $u_1 = 0.01$ and $u_2 = 0.01$.
doi:10.1371/journal.pone.0065724.g002

$R = 0$), we found two parameter regions for stochastic tunneling (i) when the fitness of type-1 cells is smaller than that of type-0 cells and fitness of type-2 is the largest and (ii) when the fitness of type-1 cells is larger than that of type-0 and the fitness of type-2 cells is slightly smaller than that of type-1 cells (**Figure 5a**). In the figure, the non-red region represents bad fit (more than 20% overestimation or underestimation) of approximations against simulation results. We also showed the comparison between (i) the simulation results and tunneling formula in the previous paper (Eq. 1 in [22]), and (ii) the simulation results and our new formula (Eq. 15) (**Figure 5b–c**). In the region where $r_1 > 1$ and $r_2 \approx r_1$, the new approximation works better than the previous formula, but when $r_1 > r_2$ the new one underestimates the simulation results and the old one fits the simulation results more accurately.

Finally, we investigated the improvement of our new approximation over existing predictions of the fixation probability of type-2 cells at time t (**Figure 6**). First, from direct computer

simulations and for each parameter set, we obtained the time at which the fixation probability of type-2 cells is 0.5. We then used these quantities for comparison with the predictions of existing approaches. **Figure 6** shows the predictions of our formula (Eq. 15) and that by Iwasa et al. [22] (Eq. 1 and Eq. 6). We found that, when the total number of cells was small (**Figure 6a–h**), both formulas showed good predictions except when both mutation rates were large (**Figure 6a and 6e**). In that case, our new formula worked better than the previous one [22]. When the total number of cells was large (**Figure 6i–p**), the previous formula did not work well. Also, when both fitness values of type-1 and type-2 were larger than 1, the new formula showed a better fit than the previous one.

We also investigated the accuracy of four other published approaches: those by Komarova et al. [20] (Section 3.2), Nowak et al. [21] (Eq. 6), Weissman et al. [28] (Eq. 25) and Proulx [29] (Eq. 11) (**Figure S3**). These four formulas did not exhibit as good a fit

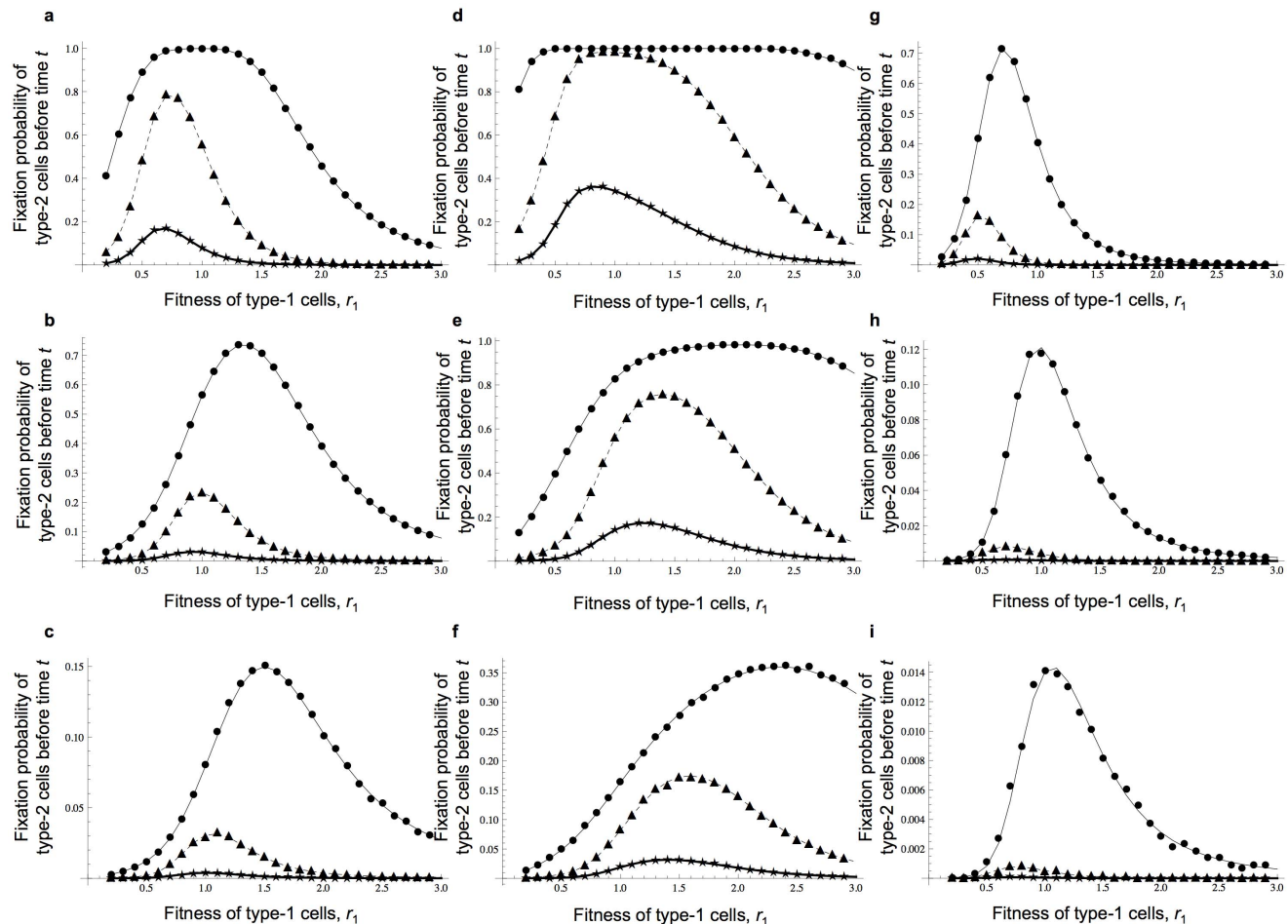


Figure 3. Precise predictions of the fixation probability of type-2 cells by systematic calculations of all transitions. The figure shows the dependence of the probability that type-2 cells are fixed at time t on various parameters. Results by systematic calculations, $W(0,0,t)$, are indicated in curves and those from direct computer simulations are shown by dots. Parameter values are $r_0 = 1$ and $t = 100$; $N = 10$; (a–c) $r_2 = 1.0$; (d–f) $r_2 = 2.0$; (g–i) $r_2 = 0.5$; (a), (d), and (g) $u_1 = 0.1$; (b), (e), and (h) $u_1 = 0.01$; and (c), (f), and (i) $u_1 = 0.001$. Circles and thin curves represent $u_2 = 0.1$, triangles and dotted lines represent $u_2 = 0.01$, and stars and bold lines represent $u_2 = 0.001$.
doi:10.1371/journal.pone.0065724.g003

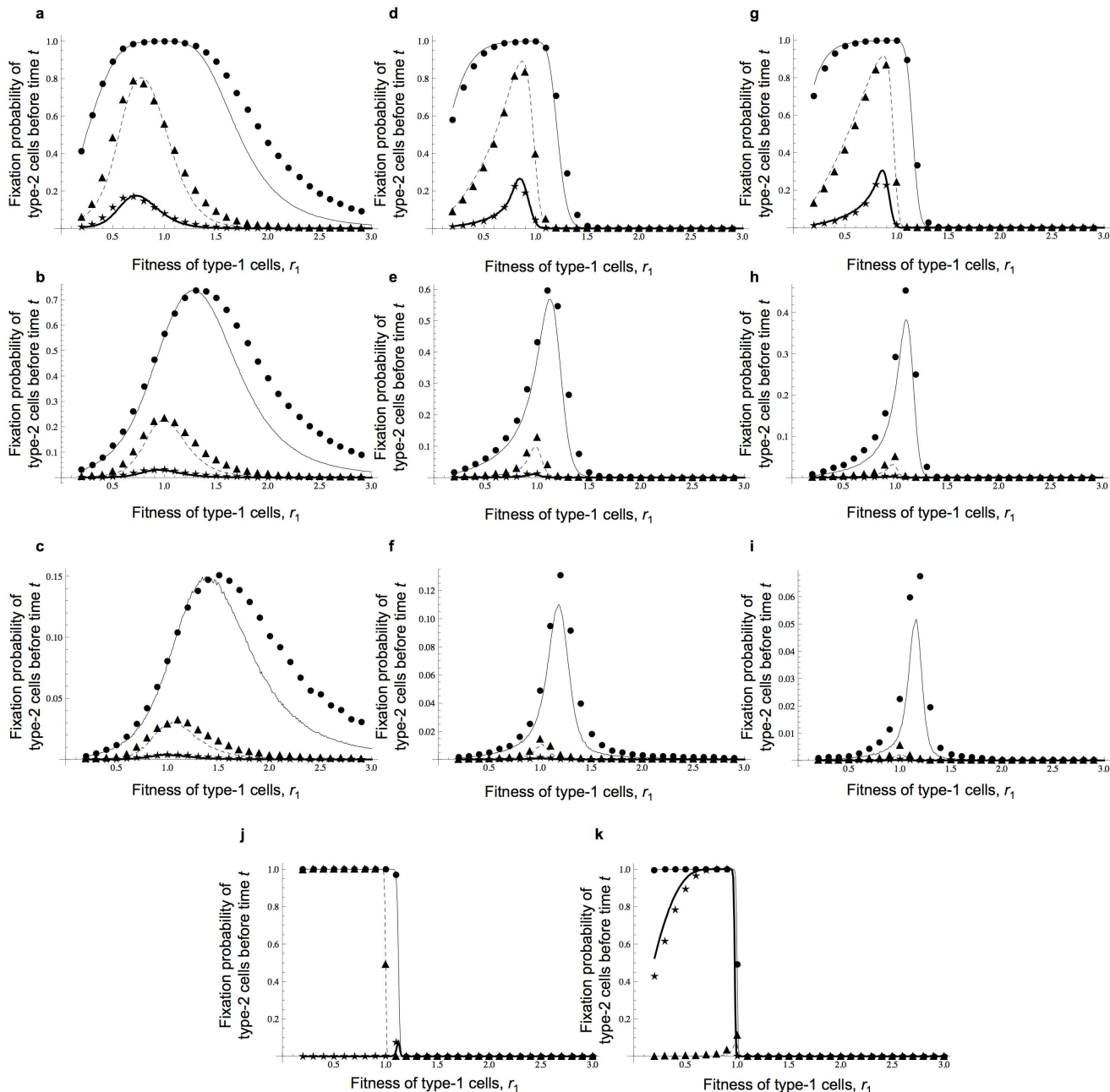


Figure 4. Results from a Wright-Fisher approximation. The figure shows the dependence of the probability that type-2 cells are fixed at time t on various parameters. Results by a Wright-Fisher framework are indicated by curves and those from direct computer simulations are shown by dots. Parameter values are $r_0 = r_2 = 1$ and $t = 100$; (a–c) $N = 10$; (d–f) $N = 50$; (g–i) $N = 100$; (a), (d), and (g) $u_1 = 0.1$; (b), (e), and (h) $u_1 = 0.01$; and (c), (f), and (i) $u_1 = 0.001$. (a–i) Circles and thin curves represent $u_2 = 0.1$, triangles and dotted lines represent $u_2 = 0.01$, and stars and bold lines represent $u_2 = 0.001$. (j and k) $N = 1000$; (j) circles and thin curves represent $u_1 = 0.0001$ and $u_2 = 0.1$, triangles and dotted lines represent $u_1 = 0.0001$ and $u_2 = 0.001$; (k) circles and thin curves represent $u_1 = 0.0001$ and $u_2 = 0.01$, triangles and dotted lines represent $u_1 = 0.0001$ and $u_2 = 0.0001$. doi:10.1371/journal.pone.0065724.g004

against the results obtained by direct computer simulations as our work. Since Komarova et al. [20] considered the probability of the first appearance of type-2 cells at time t , and not the fixation of type-2 cells, their equation always overestimates the probability of fixation (**Figure S3a–h**). The predictions by Nowak et al. [21] displayed a good fit in a certain parameter region (**Figure S3i–p**). Especially when the total cell number is large and the fitness of type-1 cells is smaller than that of type-0 cells, their predictions fit the results by direct computer simulations for a certain range of

mutation rates. Overall, however, their predictions did not work well because they also did not consider the fixation of type-2 cells, but their appearance. Weissman et al. [28] performed a comprehensive study of the tunneling rate and the expected time until a mutant with k mutations appears in asexually reproducing populations. The case of $k = 2$ represents the same condition as in our current study. In their paper [28], the authors show the tunneling rate as Eq. 25; we included their tunneling rate in the tunneling term, b , in Eq. 1 in our formula ($b = Nu_1p_1$, where p_1 is

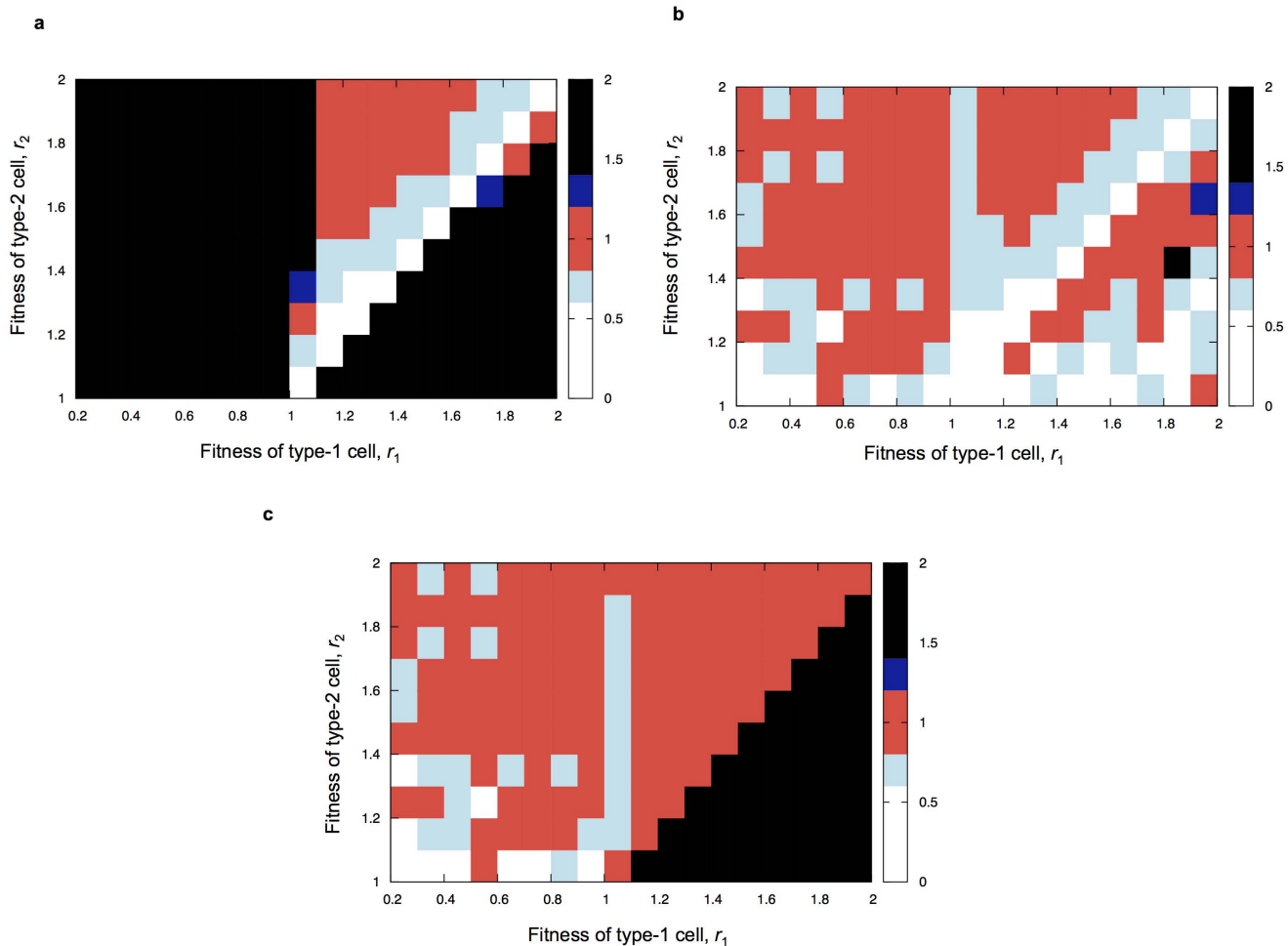


Figure 5. The region of tunneling. The figure shows a comparison between the simulation results and (panel a) Eq. (1) in reference [22] with $R=0$, (panel b) the tunneling formula in the previous paper (Eq. 1 in [22]), and (panel c) our new formula (Eq. 15). The color represents the fit between simulation results and each formula. The deviation is within 20% in a red region, from 20% to 40% overestimation in a light blue region, from 20% to 40% underestimation in a dark blue region, more than 40% overestimation in a white region, and more than 40% underestimation in a black region. Parameter values are $r_0 = 1$, $t = 200$, $N = 100$, $u_1 = 5 \cdot 10^{-4}$ and $u_2 = 10^{-3}$. doi:10.1371/journal.pone.0065724.g005

given by Eq. 25 in [28]). We found that their approach did not provide a very accurate fit to results obtained from exact computer simulations (**Figure S3q–x**). Finally, the predictions by Proulx also did not exhibit a good fit (**Figure S3y–F**).

In summary, our new approach (Eq. 15) displayed the best fit against the direct computer simulations amongst all formulas investigated (**Figure 6**). However, we still need to perform systematic calculations of all transitions (Eq. 19) or direct computer simulations when the mutation rates and the population size are large and the fitness of type-1 is close to that of type-0 cells ($r_1 \approx r_0$) (**Figure 6a and i–l**). Moreover, the new formula does not work well when (i) type-1 cells are disadvantageous and the mutation rates and the population size are large (**Figure 6i**); and when (ii) type-1 cells are advantageous, mutation rates are small and the population size is large (**Figure 6l**). In such parameter regions, the systematic calculation of all transitions (Eq. 19) or direct computer simulations are necessary.

Discussion

In this paper, we have performed a comprehensive analysis of the fixation probability of cells harboring two mutations; these

mutations are accumulated sequentially in cells within a population of fixed size. Although the evolutionary dynamics of cells acquiring one or two mutations has been studied for decades, this work represents the first investigation of the fixation probabilities in the Moran model in a wide parameter region including large mutation rates and a disadvantageous fitness of cells harboring both mutations. A consideration of the risk of a cell population harboring two mutations, as well as the fixation probability of such cells, is important for situations arising during tumorigenesis such as the inactivation of tumor suppressor genes. When the time until fixation of type-2 cells is not negligible, the latter becomes more informative than the former. This situation occurs when the fitness of type-2 cells is not sufficiently advantageous or mutation rates are very large and it is not negligible [32,33]. Our approach considers an approximation to the tunneling rate – the rate of transition from a population consisting entirely of type-0 cells to a population consisting entirely of type-2 cells, which represents an extension of our previous study [22]. This approach is computationally less expensive and provided good predictions for situations in which type-2 cells are advantageous as compared to type-0 cells. Note that we used large mutation rates in our analyses, at a range of $10^{-4} < u < 10^{-1}$; this choice was made since experimental

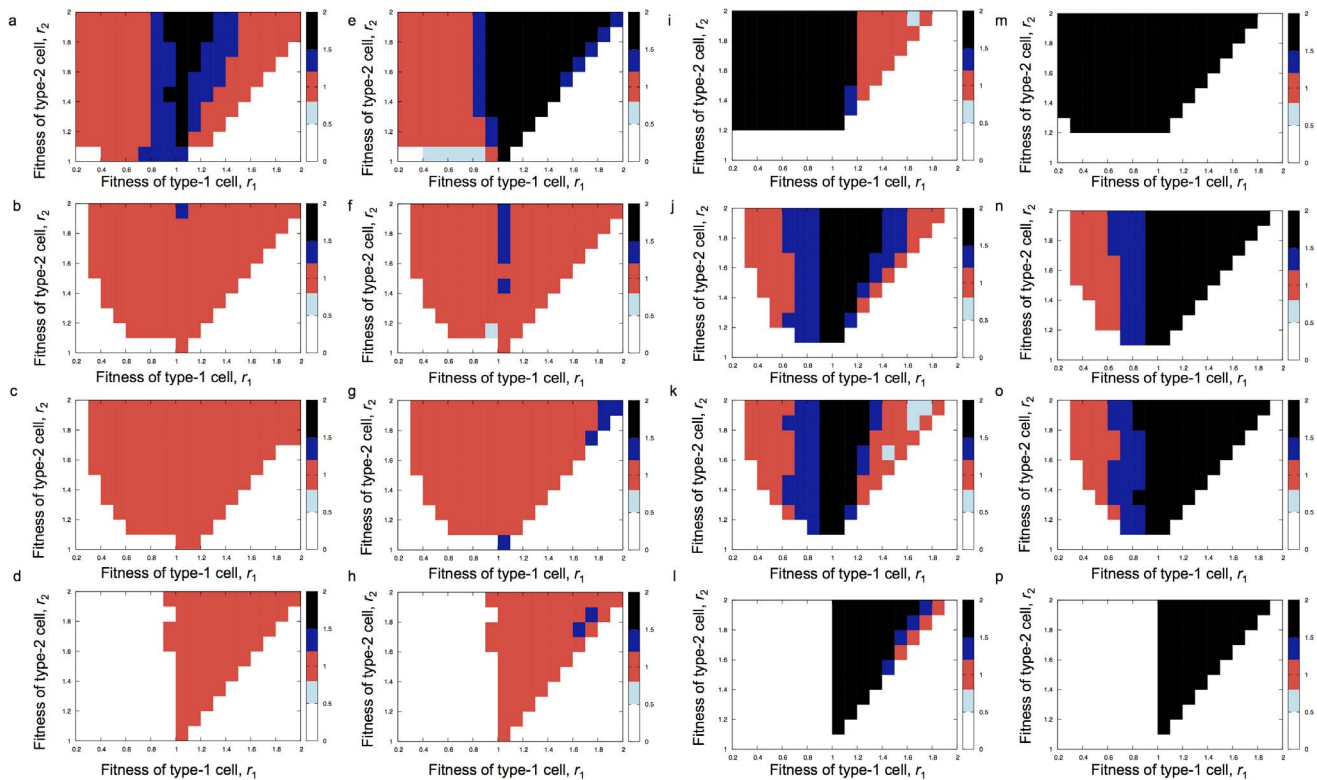


Figure 6. Predictions of the fixation probability of type-2 cells by different approaches. All predictions were divided by 0.5. When the ratio between prediction and 0.5 is 1, the color is red and signifies an accurate fit between the formula and the simulation result. When the ratio is much larger than 1, the color is blue and black and represents an overestimation of the formula. When the ratio is much smaller than 1, the color is light blue and signifies an underestimation of the formula. In the white region, we did not investigate the accuracy of the formulas because the time for the type-2 cell fixation became too long. The predictions by our approach (Eq. 15) are shown in panel a–d and i–l, and those by Iwasa et al. [22] are shown in panel e–h and m–p. Parameter values are $r_0 = 1.0$, (a–h) $N = 100$; (i–p) $N = 10000$; (a,e,i,m) $u_1 = u_2 = 0.01$; (b,f,j,n) $u_1 = 0.01$, $u_2 = 0.0001$; (c,g,k,o) $u_1 = 0.0001$, $u_2 = 0.01$; and (d,h,l,p) $u_1 = u_2 = 0.0001$. Time was chosen to obtain a probability of type-2 fixation of 0.5 for each parameter set. doi:10.1371/journal.pone.0065724.g006

evidence suggests that the mutation rate per base per cell division could increase up to these values due to phenomena such as chromosomal instability and microsatellite instability [33,34].

We then investigated an approach considering all possible states of the population, consisting of all three cell types, and calculating all transitions among these states provided accurate predictions as tested by direct computer simulations. However, the time to calculate these predictions increases as the population size expands because the number of equations increases in a factorial way, and this method is thus infeasible to perform for large populations. For situations in which type-2 cells are neutral as compared to type-0 cells, we adopted the Wright-Fisher framework to obtain computationally faster approximations of the results from the Moran model because the first method does not provide accurate predictions in this parameter region.

These results are useful for considering the dynamics of mutation accumulation during cancer initiation, progression, and the emergence of resistance. A detailed kinetic understanding of the processes leading to cells that harbor a certain number of mutations can provide greater insights into tumorigenesis as well as allow predictions for the mutational composition of a tumor at certain time points. Furthermore, such a theory allows for a study of the circumstances that maximize the rate of evolution, i.e. the rate of mutation accumulation in cell populations. When investigating the optimum fitness of type-1 cells that maximizes the probability that a cell with two mutations has reached fixation within a population of cells, we found that, in a wide parameter

region, the optimal fitness of type-1 cells is disadvantageous as compared to type-0 and type-2 cells when the first mutation rate is very large (Figure 2a, d, g). When the first mutation rate is large, then a large fraction of type-0 cell divisions contributes to an increase of type-1 cells by mutational events. This phenomenon could arise because once a mutated cell (according to our notation, a type-1 cell) appears, the clone it produces needs to undergo a large number of cell divisions to reach fixation (i.e. 100% frequency) in the population. This number of cell divisions equals at least the size of total population. During these cell divisions, the non-mutated cells (type-0 cells) experience a much larger number of cell divisions than type-1 cells because the initial number of type-0 cells is much larger than that of type-1 cells when the latter has just been produced. Then, during these cell divisions, additional mutations can emerge and will contribute to the increase of the number of type-1 cells.

Moreover several biological observations support the existence of this phenomenon. It is well known that genetic instability contributes to tumorigenesis; the rate of chromosomal loss or gain in genetically unstable cells has been measured to be about 0.01 per cell division [33]. Furthermore, mutations at different loci could result in the same phenotype of a new mutant because these mutations may affect the same signaling pathway in the cell [40], which leads to a high mutation rate for generating a particular phenotype. Finally, epigenetic changes may also occur at the same sites as genetic mutations, thus increasing the rates of alterations of

that particular locus. Similarly to genetic instability, epigenetic instability can thus also lead to large mutation rates [41].

Therefore, the relative fitness of type-1 cells as compared to type-0 cells is effectively advantageous due to the high mutation rate, even though the numerical value of type-1 cell fitness is smaller than type-0 cell fitness. Once type-1 cells become dominant in the population, a small fitness value of type-1 cells maximizes the chance of type-2 cells to reach fixation in a population of type-1 cells. Moreover, when the second mutation rate is very large, the optimal fitness of type-1 cells is even larger than that of type-2 cells (**Figure 2d–f**). A large fitness value of type-1 cells enhances the increase of type-1 cells in a population of type-0 cells. In the process of reaching fixation of type-2 cells in a population of type-1 cells, a large mutation rate enhances the abundance of type-2 cells. These phenomena could arise due to the reasons described in the previous paragraph. However, the effects of additional mutations arising during the fixation process, i.e. while the cell population increases in abundance, are thus very important for the dynamics of mutation accumulation in a population of cells; such effects have not previously been described in detail.

Moreover, we investigated parameter regions in which “stochastic tunneling” occurs. In regimes in which $r_1 < 1$ and $r_2 > 1$, the equation without a tunneling term underestimates the simulation results (**Figure 5a**). Interestingly, we found that even when $r_1 > 1$ and $r_2 < r_1$, tunneling occurs, which had not been considered as being part of the tunneling regime previously. We confirmed that when $r_1 > 1$ and $r_1 \approx r_2$, the new approximation provides more accurate predictions than the previous formulas, but when the fitness of type-2 is sufficiently smaller than that of type-1, the previous approach is better (**Figure 5b and c**).

In human tumors, the total number of cells is expected to be much larger than 1,000 cells, which is the parameter used for most of our studies. The reasons we considered relatively small population sizes are as follows: (i) Our model is not necessarily meant to consider only large, late-stage tumors with population sizes of the order of 10^9 to 10^{12} cells. The model is thus designed to describe small, constant-size populations in which sequential mutations arise, and we describe the dynamics with which this process occurs. There are only few estimates for the population structure and cell numbers within healthy human tissues; for instance, a crypt in the human colon contains about 2,000 cells, which are replenished by a small number of stem cells [4–6] [42]. Since only the mutations arising in stem cells can be maintained indefinitely within the tissue without being “washed out” of the system by differentiation, it is the number of stem cells that represents the effective population size. (ii) In many tumor types, there exists a cellular differentiation hierarchy of tumor cells, subdividing the tumor into “cancer stem cells” and “cancer differentiated cells” [43]. It has been estimated that only about one in a million tumor cells are true stem cells in tumor types that adhere to this model [44]. Since only those tumor stem cells have unlimited self-renewal capacity, the accumulation of mutations needs to be considered in only this population to study the evolutionary dynamics of the entire tumor. For those reasons and for the computational speed of our analyses, we considered relatively small population sizes.

Our findings provide new insights into the evolutionary dynamics of cancer cells. We derived a theory of the accumulation of two mutations in a population of fixed size, and found that the frequency of mutational events determines the optimum fitness landscape for cancer cells in search of accumulating multiple mutations. Once plausible parameter values have been estimated, we are now able to obtain the fixation probability of population

with two mutations at any time point. Particularly when mutations frequently occur in a cancer cell population, such as in the presence of genomic instability [32,33], the effects of multiple mutational events during a short time interval need to be considered to obtain an accurate understanding of the dynamics of cancer cells. Although we considered a single intermediate population with one mutation, two mutations may cause two types of intermediate populations; models including such extensions will be considered in the future. Moreover, our work can be extended to investigate the emergence of a larger number of mutations, as well as a population subdivision into multiple compartments or niches; these niches may harbor cells that proliferate independently of cells in neighboring niches, or there may be migration of cells from one compartment to the next. In addition, we can consider time-dependent rates of mutation, cell division and death in addition to more complicated population structures. Such studies are ongoing and will provide further insight into the somatic evolution of cancer.

Supporting Information

Figure S1 Results of our approach. The figure shows the dependence of the probability that type-2 cells are fixed in the population of cells at time t on various parameters. Results by Eq. (15) are indicated by curves and those from direct computer simulations are shown by dots. Parameter values are $r_0 = 1.0$ and $t = 100$; (a–g) $N = 50$; (h–m) $N = 100$; (a–c, h–j) $r_2 = 1.0$; (d–f, k–m) $r_2 = 2.0$; (g) $r_2 = 0.5$; (a), (d), (g), (h), and (k) $u_1 = 0.1$; (b), (e), (i), and (l) $u_1 = 0.01$; and (c), (f), (j), and (m) $u_1 = 0.001$. Circles and thin curves represent $u_2 = 0.1$, triangles and dotted lines represent $u_2 = 0.01$, and stars and bold lines represent $u_2 = 0.001$. (TIFF)

Figure S2 Precise predictions of the fixation probability of type-2 cells by systematic calculations of all transitions. The figure shows the dependence of the probability that type-2 cells are fixed at time t on various parameters. Results by systematic calculations, $W(0,0,t)$, are indicated by curves and those from direct computer simulations are shown by dots. Parameter values are $r_0 = 1.0$ and $t = 100$; (a–g) $N = 50$; (h–m) $N = 100$; (a–c, h–j) $r_2 = 1.0$; (d–f, k–m) $r_2 = 2.0$; (g) $r_2 = 0.5$; (a), (d), (g), (h), and (k) $u_1 = 0.1$; (b), (e), (i), and (l) $u_1 = 0.01$; and (c), (f), (j), and (m) $u_1 = 0.001$. Circles and thin curves represent $u_2 = 0.1$, triangles and dotted lines represent $u_2 = 0.01$, and stars and bold lines represent $u_2 = 0.001$. (TIFF)

Figure S3 Predictions of the fixation probability of type-2 cells by different approaches. This figure shows the results obtained using different approaches to calculate the fixation probability. The parameter values were chosen such that we obtained 0.5 for the fixation probability and the predictions by the formulas were divided by 0.5. When the ratio between prediction and 0.5 is 1, the color is red and signifies an accurate fit between the formula and the simulation result. When the ratio is much larger than 1, the color is blue and black and represents an overestimation of the formula. When the ratio is much smaller than 1, the color is light blue and signifies an underestimation of the formula. In the white region, we did not investigate the accuracy of the formulas because the time for the type-2 cell fixation became too long. In panels j, l, and p, when r_1 is around 1.0, the predictions underestimate the simulation results and the white region between 0 and 0.1 appears. The predictions by the formula in Komarova et al. [20], in Nowak et al. [21], in Weissman et al. [28], and in Proulx [29] are shown in panel a–h in

i-p, in q-x and in y-F, respectively. The color scheme in panel i-p was changed in order to be able to distinguish the underestimation by the formula from the low probability of fixation. Parameter values are $r_0 = 1.0$, $N = 100$ (a-d, i-l, q-t, y-B); $N = 10000$ (e-h, m-p, u-x, C-F); $u_1 = u_2 = 0.01$ (a,e,i,m,q,u,y,C); $u_1 = 0.01$, $u_2 = 0.0001$ (b,f,j,n,r,v,z,D); $u_1 = 0.0001$, $u_2 = 0.01$ (c,g,k,o,s,-w,A,E); and $u_1 = u_2 = 0.0001$ (d,h,l,p,t,x,B,F). (TIFF)

References

- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646–674.
- Knudson AG Jr (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68: 820–823.
- Rowley JD (1973) Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 243: 290–293.
- Gishizky ML, Witte ON (1992) Initiation of deregulated growth of multipotent progenitor cells by bcr-abl in vitro. *Science* 256: 836–839.
- Feinberg AP, Tycko B (2004) The history of cancer epigenetics. *Nat Rev Cancer* 4: 143–153.
- Shah NP, Skaggs BJ, Branford S, Hughes TP, Nicoll JM, et al. (2007) Sequential ABL kinase inhibitor therapy selects for compound drug-resistant BCR-ABL mutations with altered oncogenic potency. *J Clin Invest* 117: 2562–2569.
- Nordling CO (1953) A new theory on cancer-inducing mechanism. *Br J Cancer* 7: 68–72.
- Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* 8: 1–12.
- Fisher JC (1958) Multiple-mutation theory of carcinogenesis. *Nature* 181: 651–652.
- Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, et al. (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323: 643–646.
- Michor F, Hughes TP, Iwasa Y, Branford S, Shah NP, et al. (2005) Dynamics of chronic myeloid leukaemia. *Nature* 435: 1267–1270.
- Michor F, Iwasa Y, Rajagopalan H, Lengauer C, Nowak MA (2004) Linear model of colon cancer initiation. *Cell Cycle* 3: 358–362.
- van Leeuwen IM, Byrne HM, Jensen OE, King JR (2006) Crypt dynamics and colorectal cancer: advances in mathematical modelling. *Cell Prolif* 39: 157–181.
- Roeder I, Horn M, Glauche I, Hochhaus A, Mueller MC, et al. (2006) Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications. *Nat Med* 12: 1181–1184.
- Johnston MD, Edwards CM, Bodmer WF, Maini PK, Chapman SJ (2007) Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proc Natl Acad Sci U S A* 104: 4008–4013.
- Enderling H, Chaplain MA, Anderson AR, Vaidya JS (2007) A mathematical model of breast cancer development, local treatment and recurrence. *J Theor Biol* 246: 245–259.
- Dingli D, Michor F (2006) Successful therapy must eradicate cancer stem cells. *Stem Cells* 24: 2603–2610.
- Iwasa Y, Nowak MA, Michor F (2006) Evolution of resistance during clonal expansion. *Genetics* 172: 2557–2566.
- Komarova NL, Wodarz D (2007) Effect of cellular quiescence on the success of targeted CML therapy. *PLoS One* 2: e990.
- Komarova NL, Sengupta A, Nowak MA (2003) Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *J Theor Biol* 223: 433–450.
- Nowak MA, Michor F, Komarova NL, Iwasa Y (2004) Evolutionary dynamics of tumor suppressor gene inactivation. *Proc Natl Acad Sci U S A* 101: 10635–10638.
- Iwasa Y, Michor F, Nowak MA (2004) Stochastic tunnels in evolutionary dynamics. *Genetics* 166: 1571–1579.
- Weinreich DM, Chao L (2005) Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution*, 59: 1175–82.
- Schweinsberg J (2008) The waiting time for m mutations. *Electronic Journal of Probability* 13: 1442–1478.
- Lynch M (2010) Scaling expectations for the time to establishment of complex adaptations. *Proc Natl Acad Sci U S A* 107: 16577–82.
- Weissman DB, Feldman MW, Fisher DS (2010) The rate of fitness-valley crossing in sexual populations. *Genetics* 186: 1389–1410.
- Altland A, Fischer A, Krug J, Szendro IG (2011) Rare events in population genetics: stochastic tunneling in a two-locus model with recombination. *Phys Rev Lett* 106: 088101.
- Weissman DB, Desai MM, Fisher DS, Feldman MW (2009) The rate at which asexual populations cross fitness valleys. *Theor Popul Biol* 75: 286–300.
- Proulx SR (2011) The rate of multi-step evolution in Moran and Wright-Fisher populations. *Theor Popul Biol* 80: 197–207.
- Durrett R in *Probability Models for DNA Sequence Evolution*. (2008)(ed J; Heyde Gani, C C; Jagers, P; Jurtz, T G) (Springer Science+Business Media, LLC, 2008).
- Lehmann L, Roussel F (2009) Perturbation expansions of multilocus fixation probabilities for frequency-dependent selection with applications to the Hill-Robertson effect and to the joint evolution of helping and punishment. *Theor Popul Biol* 76: 35–51.
- Loeb LA (2001) A mutator phenotype in cancer. *Cancer Res* 61: 3230–3239.
- Lengauer C, Kinzler KW, Vogelstein B (1997) Genetic instability in colorectal cancers. *Nature* 386: 623–627.
- Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* 396: 643–649.
- Perucho M (1996) Cancer of the microsatellite mutator phenotype. *Biol Chem* 377: 675–684.
- Rajagopalan H, Nowak MA, Vogelstein B, Lengauer C (2003) The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer* 3: 695–701.
- Iwasa Y, Michor F (2011) Evolutionary dynamics of intratumor heterogeneity. *PLoS One* 6: e17866.
- Moran PAP (1962) *The statistical processes of evolutionary theory*. Oxford: Clarendon Press. 200 p. p.
- Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321: 1801–1806.
- Grady WM, Carethers JM (2008) Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology* 135: 1079–1099.
- Nicolas P, Kim KM, Shibata D, Tavaré S (2007) The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Comput Biol* 3: e28.
- Reya T, Morrison SJ, Clarke MF, Weissman IL (2001) Stem cells, cancer, and cancer stem cells. *Nature* 414: 105–111.
- Quintana E, Shackleton M, Sabel MS, Fullen DR, Johnson TM, et al. (2008) Efficient tumour formation by single human melanoma cells. *Nature* 456: 593–598.

Acknowledgments

The authors would like to thank the Michor lab for helpful discussions.

Author Contributions

Conceived and designed the experiments: HH YI FM. Performed the experiments: HH YEM YI FM. Analyzed the data: HH YEM YI FM. Contributed reagents/materials/analysis tools: HH YEM YI FM. Wrote the paper: HH YEM YI FM.